AIP | The Journal of Chemical Physics

Centrality measures highlight proton traps and access points to proton highways in kinetic Monte Carlo trajectories

Rachel A. Krueger, Frederick G. Haibach, Dana L. Fry, and Maria A. Gomez

Citation: The Journal of Chemical Physics **142**, 154110 (2015); doi: 10.1063/1.4917469
View online: http://dx.doi.org/10.1063/1.4917469
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/15?ver=pdfcov
Published by the AIP Publishing

Centrality measures highlight proton traps and access points to proton highways in kinetic Monte Carlo trajectories

Rachel A. Krueger,[1] Frederick G. Haibach,[2] Dana L. Fry,[3] and Maria A. Gomez[3,a)]
[1]*Department of Chemistry, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Confluent Science, Wilbraham, Massachusetts 01095, USA*
[3]*Department of Chemistry, Mount Holyoke College, South Hadley, Massachusetts 01075, USA*

(Received 12 February 2015; accepted 1 April 2015; published online 17 April 2015)

A centrality measure based on the time of first returns rather than the number of steps is developed and applied to finding proton traps and access points to proton highways in the doped perovskite oxides: $AZr_{0.875}D_{0.125}O_3$, where A is Ba or Sr and the dopant D is Y or Al. The high centrality region near the dopant is wider in the $SrZrO_3$ systems than the $BaZrO_3$ systems. In the aluminum-doped systems, a region of intermediate centrality (secondary region) is found in a plane away from the dopant. Kinetic Monte Carlo (kMC) trajectories show that this secondary region is an entry to fast conduction planes in the aluminum-doped systems in contrast to the highest centrality area near the dopant trap. The yttrium-doped systems do not show this secondary region because the fast conduction routes are in the same plane as the dopant and hence already in the high centrality trapped area. This centrality measure complements kMC by highlighting key areas in trajectories. The limiting activation barriers found via kMC are in very good agreement with experiments and related to the barriers to escape dopant traps. [http://dx.doi.org/10.1063/1.4917469]

I. INTRODUCTION

Graph theory has long been used to study a variety of networks, including social, telecommunications, ecological, cellular signaling, and citation networks. To apply the tools of graph theory to the analysis of chemical configurations, such as those generated using molecular simulation techniques, chemical species may be represented as vertices in graphs, with interactions between species represented by the graph edges. Measures of the relative importance or "centrality" of vertices[1–3] in chemical graphs often play a key role in this analysis. Software algorithms developed to investigate structural patterns and changes in hydrogen-bonded solvents[4] and protein structure[5] use the concept of degree centrality. PageRank[6] centrality has been used to describe solvent shell organization.[7]

In this contribution, centrality measures are used to gain insight into ion conduction in solids. Mechanisms of conduction and activation barriers for transport of protons[8–10] and other ions[11,12] are sometimes determined based on conduction pathways built by concatenating single-step ion transfers between sites. This is true for many solid state conductors. A wider range of conduction pathways may also be extracted from trajectories generated using kinetic simulation schemes such as the kinetic Monte Carlo (kMC) algorithm.[13] A global, qualitative description of key areas for ion flow, traps or key nexuses, would be invaluable, allowing clearer characterization and comparisons between systems. We develop a centrality measure based on time of first returns which highlights the key proton conduction regions in doped perovskites. Doped perovskites have attracted interest due to their possible use in proton conducting fuel cells.[14–16] Our study of proton conduction in doped perovskites provides a framework for similar centrality analysis of other solid systems.

Our earlier work[17–20] showed how long-range proton conduction pathways can be found in perovskites by searching a conduction graph. The graph is a collection of vertices (proton binding sites) connected by edges when there is a single transition state between the two proton binding sites. The edges can be weighted by a function of the rate constant for proton motion from one binding site to the other. The binding sites and transition states found were in qualitative agreement with those found in other work[21–25] for these and related perovskite systems. Our studies found all possible sites and made use of the whole network or graph to find conduction pathways. Similar graphs may be defined for any system where distinct binding sites and rate constants for transfer between them may be determined.

There are several ways to think about centrality of vertices in a graph.[1–3] For many applications in ergodic systems, the average number of round trip steps between vertex i and any other vertex is shorter for more central vertices. Hence, the inverse of the average number of round trip steps between vertex i and any other vertex is one reasonable measure of the centrality of vertex i. Alternatively, the difference between the average number of steps connecting any two vertices when connecting paths must pass through a vertex i versus not passing through vertex i may be calculated. The inverse of this difference is another measure of centrality of i. Both of these potential centrality measures rely on knowing the number of round trip steps between two vertices. Several theorems in

a)Electronic mail: magomez@mtholyoke.edu

Grinstead and Snell[3] show that the mean number of steps for first arrival at a vertex j starting from a vertex i can be found by diagonalizing a fundamental matrix.

In physical systems with predetermined rate constants, it is desirable to replace the mean number of steps with mean transit time. The mean transit time is more physically meaningful since different single steps from one vertex to another will have different barriers and attempt frequencies and hence have different average transit times and rate constants. Comparison with a variety of dynamic simulation methods with varying time length steps will also be facilitated by changing from mean number of steps to mean transit time. Better conductors allow ions to move quickly through the network of conduction paths. Section II of this contribution describes an approach similar to Grinstead and Snell[3] for mean time rather than mean number of steps. The needed theorems are proven in the Appendix. Using these theorems, centrality measures based on time rather than number of steps are calculated. Section II B describes the calculations needed to form the graphs for proton binding sites in four doped perovskite systems: $AZr_{0.875}D_{0.125}O_3$, where A is Ba or Sr and the dopant D is Y or Al. Sec. III shows how a visual global centrality view provides similar information to kinetic Monte Carlo. The information is not available from other measures considered. The resulting centrality measure adds to a comprehensive understanding of proton conduction in perovskite oxides as described in Sec. IV.

II. AVERAGE TIME OF FIRST RETURN AND CENTRALITY

In this application, we will define centrality of a vertex as the inverse of the average time of first return to a vertex after going to any other vertex. The average time of first return to a vertex i after going to any other vertex can be obtained by adding the mean first passage time to go from vertex i to j and the mean first passage time to go from vertex j to i and averaging over all possible vertices j. The mean first passage time to go from vertex i to j is the mean time to first get to j from i. If the average time of first return to a vertex is small, then the vertex is easily reachable from other vertices and has high centrality in the graph. This measure highlights traps and nexuses but does not directly reveal conduction pathways. Binding sites with the shortest average times of first return are most likely to be traps with high centrality. Sites that connect long-range conduction pathways to traps should have longer average times of first return or lower centrality, and sites not connected to traps likely have the highest average times of first return or lowest centrality. Overall, the centrality measure should highlight sites that reduce the capability of the material to conduct protons, with most central sites being traps and another category of central sites being connections to traps.

The average time of first returns can be found from the mean first passage time. Grinstead and Snell[3,26] highlight a variety of theorems for finding the mean first passage number of steps (n_{ij}) in ergodic Markov chains. Mean first passage number of steps from i to j in their text is defined as the expected number of steps to reach j from i for the first time. These ideas have been used by several researchers including Zhang *et al.*[1] and White and Smyth[2] to find round trip mean number of steps to go from vertex i to j for the first time and back to i for the first time since visiting j. The most straightforward calculation requires adding n_{ij} to n_{ji}. Averaging over all j gives the mean round trip time from i to any j and back to i for the first time or the mean number of steps to first return to i after going to any other vertex. The smaller the average number of steps to first return to i after going to any other vertex, the more central i is. Hence, a common centrality measure at i is the inverse of this average.

In the Appendix, we develop how to obtain the average time of first return to a vertex after going to any other vertex by closely following Grinstead and Snell's approach,[3] except that instead of considering number of steps, we will consider the average time for same set of steps. The time to go from i to j is just the inverse of the rate constant, k_{ij}. This approach leads to the mean round trip passage time to go from i to j and back to i or the average time of first return to i after going to j for the first time or Eq. (1) here and Eq. (A7) in the Appendix,

$$R_{ij} = \left(\frac{Z_{jj} - Z_{ij}}{\pi_j} + \frac{Z_{ii} - Z_{ji}}{\pi_i}\right) \sum_n \pi_n c_n. \tag{1}$$

$\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1}$ and is the fundamental matrix for ergodic chains[3] or the key matrix in deriving a variety of properties about ergodic chains. $\mathbf{I}$ is the identity matrix. $\mathbf{P}$ is the matrix of probabilities of going from site i to site j. $\mathbf{W}$ is a matrix whose rows are $\boldsymbol{\pi}$. π_i is the stationary probability for site i. c_n is the expected time for a first step starting at n.

A. Quantities needed to calculate mean round trip passage times

Evaluation of Eq. (1) requires knowing the fundamental matrix ($\mathbf{Z}$), the stationary probability for all sites i (π_i), and the expected time for the first step starting at any n (c_n). Here, we outline the elements needed for the calculation as well as the data required to find these elements.

1. Preliminaries obtained using a combination of *ab initio*, conjugate gradient minimization, transition state finding methods, and normal mode analysis.
   (a) Proton binding site energies (E_i) and normal mode frequencies ($\nu_k^{\mathrm{MIN}_i}$) for vertex i and mode k.
   (b) Transition state energies (E_{ij}^{TS}) and normal mode frequencies ($\nu_k^{\mathrm{TS}_{ij}}$) between adjacent vertices i and j and mode k.
   (c) Rate constants: We use harmonic transition state theory constants given by

$$k_{ij}^{\mathrm{TS}} = \frac{\Pi_{k=1}^{M} \nu_k^{\mathrm{MIN}_i}}{\Pi_{k=1}^{M-1} \nu_k^{\mathrm{TS}_{ij}}} \exp(-\beta(E_{ij}^{\mathrm{TS}} - E_i)).$$

   The probability to go from i to j is a single step which is given by the normalized rate constant for that step, i.e., $p_{ij} = \frac{k_{ij}}{\sum_n k_{in}}$.
2. Elements in Eq. (A7) are as follows.
   (a) *Stationary probability* π_j: When choosing the stationary probability at site j, π_j, it is necessary to ensure

that the probability satisfies detailed balance.[13] When harmonic transition state theory rate constants are used, the detailed balance condition $\pi_i k_{ij}^{TS} = \pi_j k_{ji}^{TS}$ is satisfied with $\pi_i = \frac{\exp(-\beta E_i)}{\prod_{k=1}^{M} \nu_k^{MIN_i} Q}$, where Q is the normalization factor. This is the Boltzmann distribution for the potential of the electronic energy plus the normal mode harmonic vibrational corrections integrated over all normal modes. We sample from the *NVT* ensemble.

(b) *Fundamental matrix* **Z**: The fundamental matrix for ergodic chains is given by $(\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1}$, where **I** is the identity matrix, **P** is the transition probability matrix, and **W** is a matrix whose rows are $\boldsymbol{\pi}$.[3] The probabilities in the transition probability matrix ($\mathbf{P}_{ij} = p_{ij}$) are simply the normalized rate constants $\frac{k_{ij}^{TS}}{\sum_j k_{ij}^{TS}}$. The fundamental matrix is found by inverting[27] $\mathbf{I} - \mathbf{P} + \mathbf{W}$.

(c) The expected time for a first step starting at n is $\sum_l p_{nl} \frac{1}{k_{nl}}$ or the average of the inverse of the rate constant starting at n and ending at any other vertex connected to n by a single transition state. The probability of going from n to l is the normalized rate constant for that move ($p_{nl} = \frac{k_{nl}}{\sum_j k_{nj}}$).

Together, these quantities give all the information needed to calculate the mean first passage time for going from i to j and back to i. Averaging overall possible intermediate points j gives us the average time of first returns to vertex i. The inverse of this average is our centrality. Additional details are in the Appendix.

B. Density functional calculations for quantities needed for mean round trip times

The binding site energies, transition state energies, and normal mode frequencies at transition states and binding sites are the key quantities needed to calculate the rate constant for each move between vertices in the graph. These quantities have been calculated in earlier papers[17,18,20] using the density functional theory implementation in the Vienna *Ab-Initio* Simulation Package (VASP)[28–31] with the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional within a generalized gradient approximation (GGA). The projector augment wave (PAW) method[32] was used with the valence states of Ba(5s, 5p, 6s, 5d), Sr (4s, 4p, 5s, 4d), Zr(4s, 4p, 5s, 4d), Y(4s, 4p, 5s, 4d), Al (3s, 3p), and O(2s, 2p).[33] Gaussian smearing is used to calculate partial wave occupancies. The smearing width starts at 0.2 eV and is extrapolated to zero. Each band was optimized iteratively using a preconditioned residual minimization method with direct inversion in the subspace. Projection operators were evaluated in reciprocal space. Simulation boxes of $2 \times 2 \times 2$ cells were used with a $2 \times 2 \times 2$ Monkhorst-Pack k-point mesh with zero shift. Using these conditions, all relaxed binding site energies, relaxed transition state energies, and normal modes at both binding and transition states were calculated using a conjugate gradient method, a nudged elastic band method,[34] and standard normal mode analysis,[35] respectively. For further details of the *ab initio* methods used as well as the binding and transitions states found, see Refs. 17, 18, and 20. We use these results to estimate the key quantities of the stationary distribution and harmonic transition state rate constants.

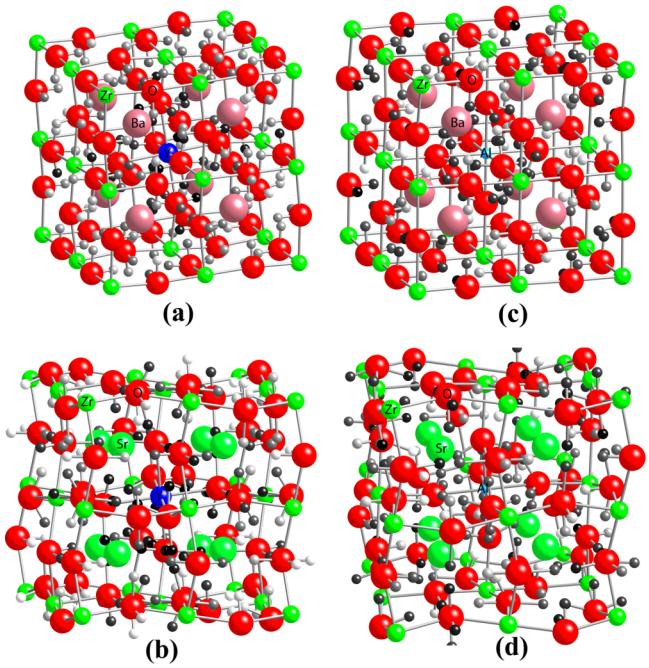

FIG. 1. The centralities for all binding sites in (a) $BaZr_{0.875}Y_{0.125}O_3$, (b) $SrZr_{0.875}Y_{0.125}O_3$, (c) $BaZr_{0.875}Al_{0.125}O_3$, and (d) $SrZr_{0.875}Al_{0.125}O_3$ at 1000 K are shown by gray scale value. Barium ions are shown in pink; strontium ions are shown as large green spheres; zirconium ions are depicted by small green spheres; yttrium ions are shown in dark blue; aluminum ions are shown in light blue; and oxygen ions are shown in red. All proton binding sites are shown in gray scale with darkest sites being the most central. In the Y doped cases (a) and (b), the most central sites (darkest) are on oxygens next to the dopant and a few are on corridors that would lead to the next dopant with the distribution of central sites being broader for the $SrZrO_3$ system showing longer range trapping. In contrast, in the Al doped cases (c) and (d), there are central sites both near the dopant and near planes further from the dopant.

III. CENTRALITY MEASURES GIVE A GLOBAL VISUAL RESULT THAT YIELDS SIMILAR INFORMATION TO KINETIC MONTE CARLO DYNAMICS

Centralities for each of the proton binding sites in the $AZr_{0.875}D_{0.125}O_3$ systems at 1000 K were calculated as described in Sec. II and shifted and scaled so that the largest centrality value is set to 100% (black) and the smallest to 0% (white). Visual comparison of the binding site centrality gray scale in various perovskites reveals the relative importance of the dopant and the conduction pathways in shaping proton flow.

Figure 1(a) shows the centrality plot for $BaZr_{0.875}Y_{0.125}O_3$. The most central sites are on oxygens next to the dopant and a few are on corridors that would lead to the next dopant plane. The centrality distribution of binding sites for $SrZr_{0.875}Y_{0.125}O_3$ shows a similar trend in Figure 1(b). However, the most central sites extend through the planes of the dopant more than in the $BaZrO_3$ case, suggesting longer range trapping. In contrast, Figures 1(c) and 1(d), which show the same perovskites doped with aluminum, show that the most central sites are near the dopant and on or near planes further from the dopant.

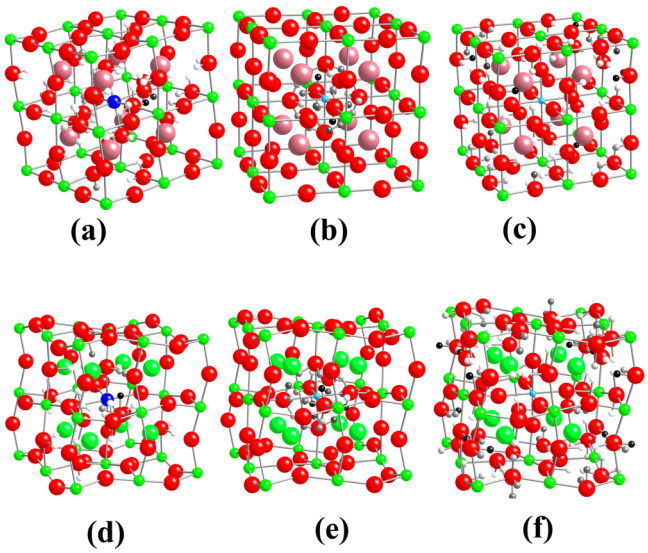

FIG. 2. Proton binding sites visited in kMC at 1000 K are shown for systems (a) $BaZr_{0.875}Y_{0.125}O_3$, (b) and (c) $BaZr_{0.875}Al_{0.125}O_3$, (d) $SrZr_{0.875}Y_{0.125}O_3$, and (e) and (f) $SrZr_{0.875}Al_{0.125}O_3$. Darker gray scale reflects that the protons have visited a site for a longer time in the simulation. In the case of the Yttrium doped systems (a) and (d), a long-range path is already seen in 16 and 31 ps, respectively. However, in the aluminum-doped systems, the proton is trapped by the dopant for 1271 ps and 260 ps in (b) and (e), respectively. After this trapping, exploration of longer range conduction planes without the dopant occurs as shown in (c) and (f). Sometimes, this results in transport through the whole simulation box, and other times, trapping occurs before long-range transport occurs.

Protons escaping from dopant traps is a potentially rare event. Molecular dynamics (MD) simulations provide a visual picture of the individual proton motions, but long simulation times are required to capture rare events because the time scale of the individual simulation steps must capture vibrational motions accurately. kMC[13] avoids the problem of waiting a long time to escape a site by using the probabilities to escape a site in all possible ways to choose a move and then advance the system clock. This allows moves of varying time duration. The information about proton vibrations around each site is lost, but in this study, the binding site long-range motion through the graph network is most important.

Figure 2(a) shows the sites visited in the first 16 ps of a kMC simulation of proton motion in $BaZr_{0.875}Y_{0.125}O_3$ at 1000 K where the proton was started at the lowest-energy binding site. Sites on the planes containing the dopant have been extensively visited in this short time range. The color shown for the sites is proportional to the time spent at the site, with the darkest colors representing the longest times spent at the site. The longest times have been spent at fairly central sites, i.e., sites right by the dopant. Further, it is apparent that planes containing dopant have been used for conduction during this short time frame. It takes about 31 ps in one trajectory to see long-range paths in $SrZr_{0.875}Y_{0.125}O_3$. Once again those long-range paths are on or near the plane containing the dopant as seen in Figure 2(d). In contrast, in Figures 2(b) and 2(e), the proton spends a significant amount of time trapped by the dopant 1271 ps and 260 ps in $BaZr_{0.875}Al_{0.125}O_3$ and $SrZr_{0.875}Al_{0.125}O_3$ at 1000 K, respectively, before exploring a plane away from the dopant where long-range conduction can occur. Figures 2(c) and 2(f) show the long-range conduction that occurs from 1271 to 1419 ps in $BaZr_{0.875}Al_{0.125}O_3$ and from 260 to 710 ps in $SrZr_{0.875}Al_{0.125}O_3$.

To calculate limiting barriers to conduction in a typical system, we would calculate the average square displacement as a function of time to extract diffusion constants. However, because there is significant trapping in these systems, the average square displacement as a function of time plot becomes non-linear before the proton escapes the trap, and hence, activation energies extracted from the diffusion constants calculated from the short linear region are just representative of the activation energy for motion within the trapped region. As an alternative, we remove the proton when it reaches one end of the simulation box and place a new proton at the starting plane. The new proton at the starting plane is placed using a Boltzmann distribution of proton binding sites at the starting plane. This is equivalent to many separate trajectories without equilibration time. While this removal breaks detailed balance and hence prevents a Boltzmann distribution of binding sites from being reached at the edges, it does allow for calculating an average limiting barrier. For each trajectory across the simulation box, a limiting barrier is logged and averaged to get the trajectory limiting barrier average. We also log what type of barrier the limiting barrier is, i.e., a rotational barrier, intra-octahedral transfer barrier, or inter-octahedral transfer barrier. With a 1 000 000 ps run with proton removals at one end of the simulation box and re-entries at the other as described above, we get the averages shown in Table I. The percent of the time that each of these barriers is encountered is also noted in Table I. Generally, the limiting barrier step is intra-octahedral transfer but sometimes it is rotation and even less often it is inter-octahedral transfer.

As Table I shows, the proton limiting barrier to long-range conduction found by kMC $BaZr_{0.875}Y_{0.125}O_3$ is 0.39 eV which is in very good agreement with the experimental barrier of 0.43 eV[36] for $BaZr_{0.9}Y_{0.1}O_3$. kMC trajectories show that most of the limiting barriers to long-range trajectories are limited

TABLE I. The perovskite, trajectory limiting barrier, and percent of the time that rotation, intra-octahedral transfer, and inter-octahedral transfer are the rate limiting step are shown for our four perovskites at 1000 K.

Perovskite	Trajectory limiting barrier (eV)	% rotation	% intra-octahedral transfer	% inter-octahedral transfer
$BaZr_{0.875}Y_{0.125}O_3$	0.39	11	87	2
$BaZr_{0.875}Al_{0.125}O_3$	0.81	0	100	0
$SrZr_{0.875}Y_{0.125}O_3$	0.57	45	55	0
$SrZr_{0.875}Al_{0.125}O_3$	0.73	12	55	33

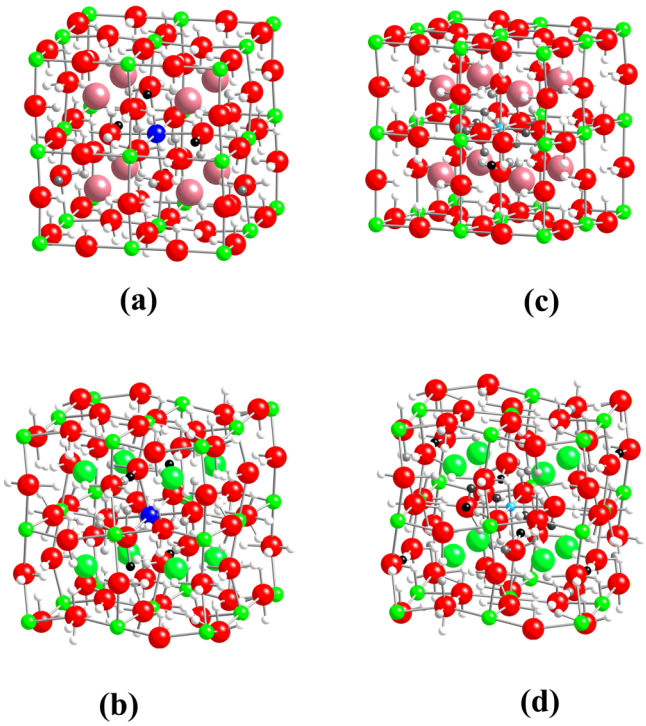

FIG. 3. Boltzmann distribution of proton binding sites for (a) $BaZr_{0.875}Y_{0.125}O_3$, (b) $SrZr_{0.875}Y_{0.125}O_3$, (c) $BaZr_{0.875}Al_{0.125}O_3$, and (d) $SrZr_{0.875}Al_{0.125}O_3$ at 1000 K. The darkest points are the most probable.

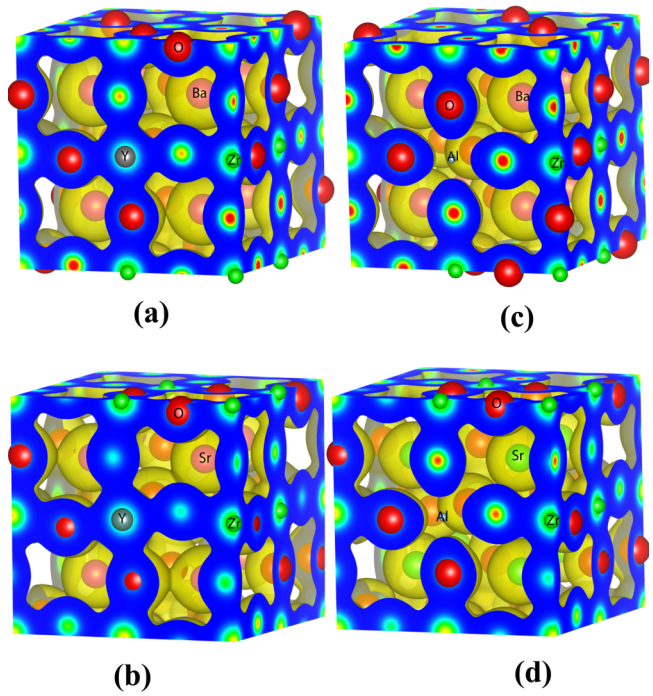

FIG. 4. Electron density contours for (a) $BaZr_{0.875}Y_{0.125}O_3$, (b) $SrZr_{0.875}Y_{0.125}O_3$, (c) $BaZr_{0.875}Al_{0.125}O_3$, and (d) $SrZr_{0.875}Al_{0.125}O_3$. The dopant is shown in gray at the face center of these unit cells for clarity. Notice that there is less electron density near the dopant for Al-doped perovskites. The pattern in electron densities appears to be similar in Ba and Sr perovskites. Plotted using VESTA.[38]

by intra-octahedral transfers. Comparison with earlier limiting barriers to periodic paths calculated by graph theory[17,19] of 0.3 eV shows a slight increase in activation energy when kMC which considers non-periodic long-range motions is used. We were not able to find an experimental study noting the activation energy for proton conduction in the corresponding aluminum-doped system. However, comparison with our earlier study[20] using a different estimation method shows a barrier range to escape dopant traps of 0.7–0.9 eV and barriers in the long-range conduction paths that the proton escapes to of about 0.4 eV. In that study, activation barriers found from diffusion constants extracted from average square displacements as a function of time showed that the linear part of this plot only captured motion within the trap and not long-range proton conduction. Motion between dopant traps is very fast and the proton is trapped at dopant sites most of the time. Removing the proton once it gets to the end of the simulation box in this study gives a long-range kMC limiting activation energy of 0.81 eV which is in the range of the energy to escape dopant traps.

Yajima *et al.*[37] show conductivity plots for 5% Y and Al doped $SrZrO_3$ in Fig. 3 of their paper. From these plots, activation energies of 0.43 eV and 0.97 eV have been extracted by Liu *et al.*[21] As shown in Table I, the kMC activation energy values for proton conduction are 0.57 eV and 0.73 eV for yttrium and aluminum doping of $SrZrO_3$ at 12.5% level. While the percent doping is different from the experiment, the trend and rough value of the activation energy is in very good agreement. The $SrZr_{0.875}Y_{0.125}O_3$ limiting barrier for periodic long-range paths of 0.4 eV[18,19] increases upon inclusion of non-periodic paths in kMC to 0.57 eV. In the aluminum-doped case, comparison with both limiting barriers for periodic long-range paths calculated via graph theory[18,19] of 0.6 eV as well as the range to escaping dopant traps of 0.7–0.8 eV found via *ab initio* methods[18] shows that proton conduction in the aluminum-doped system is limited by dopant escape and the kMC barrier is in the range of possible barriers to escape the dopant. This is in good agreement with Figure 2.

Comparing Figures 1 and 2 shows that the high centrality sites in the corners of the Al-doped systems likely provide the path needed to go from the dopant trap into the long-range conduction pathways in non-dopant planes. Notice that Figures 2(c) and 2(f) show the greatest amount of time spent on those corners and then very little time spent on the planes suggesting that the conduction is fast there. Since the greatest amount of time is spent at the traps, a quick view of the kMC trajectory would miss these important segments which the centrality measure pictures highlight as a critical area. This suggests that kMC and centrality measure pictures can complement each other, with the centrality measure picture pinpointing important areas to examine more closely in the kMC trajectory.

In contrast to centrality measures, the Boltzmann distribution and electron density do not hint that the corners farthest away from the aluminum dopant are important areas to consider. Figure 3 shows the Boltzmann distribution with black binding sites being the most probable and white the least probable. The Boltzmann distribution for all the systems mostly highlights the area near the dopant as important and does not capture the corners away from the dopant as important in the Al doped systems except very slightly in the $SrZr_{0.875}Al_{0.125}O_3$ case. Visual inspections of electron density plots of relaxed perovskites without protons show that the

oxygen atoms near the dopant have less electron density than those further away from the dopant, and in particular, the electron density by the aluminum dopant is greatly reduced, as shown in Figure 4. Naïvely, higher electron density should indicate low-energy pathways and sites for the protons. In these structures, the low electron density near the dopant and on the neighboring oxygens suggests that the proton would avoid the dopant altogether which is in direct opposition to what we see in kMC trajectories. Clearly, the charge on the proton has a significant effect on the local geometry and electrostatics. Visual inspection of the electron density in the system without the proton is not sufficiently efficient to deduce the electrostatic topology of all of the binding sites on the large number of minima and transition states needed to understand the conduction pathways. Looking at electron density plots for each binding site and transition state might be more fruitful but would no longer give a single picture with a global view. Including the effect of the proton, as occurs in the Boltzmann, kMC, and centrality figures, is necessary to reveal more details. Unlike the Boltzmann distribution and the electron density, centrality measures consider kinetic information and connectedness of sites, which allow them to give a sense of kinetically important areas in a visual way.

IV. CONCLUDING DISCUSSION

Centrality measures based on time rather than number of steps at 1000 K have been used to assess proton movement in $AZr_{0.875}D_{0.125}O_3$ perovskites, where A is Ba or Sr and the dopant D is Y or Al. This centrality measure highlights traps. This is different than looking for bound protons that occupy a single locus. Protons may be sequestered in a volume and exit that region through a nexus between the trap, other traps, and conduction pathways. The most central sites in the systems doped with yttrium are near the dopant though the width of the regions varies with the system. Further, in these yttrium-doped systems, kMC trajectories show that most of the time is indeed spent at sites by the dopant, and long-range trajectories connect dopant regions in the same plane as the dopant. In contrast, in the aluminum-doped systems, there is an additional region of high centrality in sections of planes without the dopant. kMC trajectories in the aluminum-doped systems show long time trapping near the dopant followed by rare excursions to planes without dopant where long-range conduction occurs. The areas in those planes where the proton spends more time are by the second key centrality regions found. Measures such as the Boltzmann distribution do identify the region near the dopant as a region of high probability to find protons. Visual inspection of the electron density in the absence of protons does not reveal enough details to capture the diversity of all the proton binding sites. Neither shows a second centrality region which seems to be an important area, namely, the way that a proton can escape the dopant in aluminum-doped system to access longer range conduction regions. Further, differences in the width of the trapping region are not seen in these measures.

kMC trajectories also allow for calculating of long-range diffusion limiting barriers which are in good agreement with experiment. While centrality measures do not yield these limiting barriers directly, they do suggest regions to pay attention to in trajectories and thus provide a tool complementary to kMC simulations. Both the calculation of centrality measures and kMC simulations require the same input information so centrality measures do not add much cost to a study of kMC trajectories.

ACKNOWLEDGMENTS

We would like to thank Dylan Shepardson for helpful discussions.

This research was supported by a NSF RUI Grant No. CHE-1111474 and a Cottrell Science Scholars Multi-investigator award. Computational resources were provided in part by the MERCURY supercomputer consortium http://mars.hamilton.edu under NSF MRI No. CHE-1229354.

APPENDIX: DETAILS LINKING SIMULATION STEPS TO TIME EVOLUTION

Grinstead and Snell[3,26] highlight a variety of theorems for finding the mean first passage number of steps (n_{ij}) in ergodic Markov chains. Below, we follow that approach to find mean first passage time rather than number of steps.

1. A matrix recursion equation for mean first passage time

Let m_{ij} be the mean first passage time to go from i to j. Movement from i to j occurs in either one step by going directly from i to j or by first going to another vertex l in the first step and then getting to j later on for the first time. m_{ij} is the probability of going from i to j in a single step (p_{ij}) times the average time to go from i to j in a single step ($\frac{1}{k_{ij}}$) plus the sum over the probability of going from i to the intermediate vertex, l, in the first step (p_{il}) times the combination of the mean time for the first step from i to l ($\frac{1}{k_{il}}$) and the mean first passage time to go from l to j (m_{lj}). The second equality makes use of the fact that $m_{jj} = 0$. When i is not equal to j,

$$m_{ij} = p_{ij}\frac{1}{k_{ij}} + \sum_{l\neq j} p_{il}\left(\frac{1}{k_{il}} + m_{lj}\right) = \sum_{l} p_{il}\left(\frac{1}{k_{il}} + m_{lj}\right). \tag{A1}$$

Let r_i be the mean time to first return to i after having left i or the mean recurrence time. The mean recurrence time for site i is the sum over the probability to move from i to l in one step times the sum of the time to go from i to l in one step and the mean first passage time from l to i,

$$r_i = \sum_{l} p_{il}\left(\frac{1}{k_{il}} + m_{li}\right). \tag{A2}$$

Equation (A1) can be expressed in matrix form as

$$\mathbf{M} = \mathbf{PM} + \mathbf{C} - \mathbf{D}, \tag{A3}$$

where $\mathbf{M}_{ij} = m_{ij}$, $\mathbf{P}_{ij} = p_{ij}$ or the probability of moving from i to j, $\mathbf{C}$ is a matrix with entries of $c_i \equiv c_{ij} = \sum_l \frac{p_{il}}{k_{il}}$ for every element j in the ith row, and $\mathbf{D}$ is a diagonal matrix with the mean recurrence time for site i (r_i) in the ith diagonal spot and zeros elsewhere.

2. Mean recurrence time

Theorem 11.5 from Grinstead and Snell[3] can still be recovered when the inverse rate constants are used for times rather than using number of steps, as in the original development.[3] Multiplying matrix recursion Eq. (A3) on the left by the row vector of probabilities for each state, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots \pi_N)$, and noting that moving an equilibrium population forward by one step yield the equilibrium population, i.e., $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ in an ergodic system and result in the relationship $\boldsymbol{\pi}\mathbf{C} = \boldsymbol{\pi}\mathbf{D}$. Comparing the resulting vector terms on both sides shows that the expected time for first steps starting at i is effectively an expected time for any first step divided by the probability of starting at i,

$$r_i = \frac{\sum_n \pi_n c_n}{\pi_i}, \tag{A4}$$

where $c_n = \sum_l p_{nl} \frac{1}{k_{nl}}$ is an expected time for the first step starting at n. $\sum_n \pi_n c_n$ averages these expected times starting at n by the probability of site n. The mean number of steps to first return in Grinstead and Snell[3] has a 1 in the numerator of Eq. (A4) rather than $\sum_n \pi_n c_n$ since c_n is then the expected number of steps for the first step, i.e., 1 and the sum over the probabilities for all sites times one is one. Effectively, the mean time of first return to i still remains inversely proportional to the probability of i but the proportionality constant, mean time for first step, is now carrying the units of time.

3. Some pieces that remain unchained

Grinstead and Snell[3,26] show that for an ergodic Markov chain, the π_i are strictly positive. This is true in this case as π_i is the Boltzmann distribution for site i and all sites are accessible. Further, Proposition 11.1 of Grinstead and Snell[3] is still valid. It states that if $\mathbf{P}$ is the matrix of individual step probabilities for an ergodic chain and $\mathbf{W}$ is a matrix whose rows are $\boldsymbol{\pi}$, then $\mathbf{I} - \mathbf{P} + \mathbf{W}$ has an inverse. They further show that this inverse is the fundamental matrix for ergodic Markov chains, $\mathbf{Z}$, or the key matrix to finding many properties of ergodic Markov chains. The argument for Proposition 11.1 in Grinstead and Snell[3] does not involve properties of ergodic Markov chains such as the mean first passage number of steps to go from i to j which we have changed to a mean first passage time. Hence, we use their proposition without alteration.

Parts 2 and 3 of the Lemma 11.2 in Grinstead and Snell[3] are also unchanged and so we still use the results $\boldsymbol{\pi}Z = \boldsymbol{\pi}$ and $\mathbf{Z}(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{W}$, respectively. Part 1, which finds that $\mathbf{Zc} = \mathbf{c}$, is impacted because the $\mathbf{c}$ vector is no longer a vector of ones but instead a vector with $\sum_l \frac{p_{il}}{k_{il}}$ entries for the ith element. Hence, now multiplying $\mathbf{c}$ on the left by one in the form of the fundamental matrix times its inverse does not lead to much simplification but to an identity relationship.

4. The first mean passage time matrix can be found from the fundamental matrix

Theorem 11.6 in Grinstead and Snell[3] is now modified slightly as follows: The first mean passage time matrix elements for an ergodic system are given by $m_{ij} = \frac{Z_{jj} - Z_{ij}}{\pi_j} \sum_n c_n \pi_n + \sum_n (Z_{in} - Z_{jn}) c_n$, where $\sum_n c_n \pi_n$ is the expected time for any first step and c_n is the expected time for a first step starting at n or $\sum_l p_{nl} \frac{1}{k_{nl}}$. Old Theorem 11.5 simply stated that $m_{ij} = \frac{Z_{jj} - Z_{ij}}{\pi_j}$. This is because the expected number of steps rather than time for any first step is 1 and $\mathbf{Zc} = \mathbf{c}$ (old Lemma 11.2 part 1). Now, let us prove the new theorem which considers time rather than number of steps.

Starting from Eq. (A3) with both $\mathbf{M}$ terms moved to the left-hand side, multiplying both sides by the fundamental matrix on the left, and using the third part of Lemma 11.2 ($\mathbf{Z}(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{W}$) to simplify yields

$$\mathbf{M} = \mathbf{ZC} - \mathbf{ZD} + \mathbf{WM}. \tag{A5}$$

Inserting the matrix element definitions for $\mathbf{D}$, $\mathbf{C}$, and $\mathbf{W}$ yields $m_{ij} = \sum_n Z_{in} c_n - Z_{ij} r_j + (\boldsymbol{\pi}\mathrm{M})_j$, where the meaning of $(\boldsymbol{\pi}\mathrm{M})_j$ can be found by considering the case where $i = j$, which implies a zero time ($m_{jj} = 0 = \sum_n Z_{jn} c_n - Z_{jj} r_j + (\boldsymbol{\pi}\mathrm{M})_j$) and yields $(\boldsymbol{\pi}\mathrm{M})_j = Z_{jj} r_j - \sum_n Z_{jn} c_n$. Using this relationship transforms the m_{ij} equation to $m_{ij} = \sum_n (Z_{in} - Z_{jn}) c_n + (Z_{jj} - Z_{ij}) r_j$. Finally, inserting mean recurrence time Eq. (A4) and recognizing that $c_n = \sum_l \frac{p_{nl}}{k_{nl}}$ yield our final expression for the mean first passage time,

$$m_{ij} = \frac{Z_{jj} - Z_{ij}}{\pi_j} \sum_n \pi_n c_n + \sum_n (Z_{in} - Z_{jn}) c_n. \tag{A6}$$

Notice that $\sum_n \pi_n c_n$ is just the expected time for the first step and c_n is the expected time for a first step starting at n. When the number of steps is considered rather than the time, both of these terms are 1 and original Theorem 11.6 in Grinstead and Snell[3] is recovered.

5. Round trip times and centrality measures

Now, the round trip time (R_{ij}) to first go from i to j and come back to i from j is simply $m_{ij} + m_{ji}$. Inserting Eq. (A6) and simplifying yields

$$R_{ij} = \left(\frac{Z_{jj} - Z_{ij}}{\pi_j} + \frac{Z_{ii} - Z_{ji}}{\pi_i} \right) \sum_n \pi_n c_n. \tag{A7}$$

Averaging over all the possible intermediate points j in the round trip time above tells us the average round trip time to first go from i to any j and back or the average time of first return to a vertex after going to any other vertex. The smaller the average, the more central the site i is. Hence, the inverse of this average can be thought of as a centrality measure. In our study, we set the maximum centrality found to black and create a gray scale with white as the least central point.

[1]D. Boley, G. Ranjan, and Z.-L. Zhang, Linear Algebra Its Appl. **435**, 225 (2011).
[2]S. White and P. Smyth, *Proceedings of the Night ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, 2003), pp. 266–275.
[3]C. M. Grinstead and J. L. Snell, *Introduction to Probability*, 2nd Revised ed. (American Mathematical Society, Providence, RI, 1997).
[4]A. Ozkanlar and A. E. Clark, J. Comput. Chem. **35**, 495 (2014).
[5]M. Tiberti, G. Invernizzi, M. Lambrughi, Y. Inbar, G. Schreiber, and E. Papaleo, J. Chem. Inf. Model. **54**, 1537 (2014).

[6]S. Brin and L. Page, Comput. Networks ISDN Syst. **30**, 107 (1998).
[7]B. L. Mooney, L. R. Corrales, and A. E. Clark, J. Comput. Chem. **33**, 853 (2012).
[8]K. Kato, K. Toyoura, A. Nakamura, and K. Matsunaga, J. Phys. Chem. C **118**, 9377 (2014).
[9]Y.-C. Jeong, B.-K. Kim, and Y.-C. Kim, Solid State Ionics **259**, 1 (2014).
[10]D.-H. Kim, B.-K. Kim, and Y.-C. Kim, Solid State Ionics **213**, 18 (2012).
[11]E. E. Jay, P. M. Mallinson, S. K. Fong, B. L. Metcalfe, and R. W. Grimes, J. Mater. Sci. **46**, 7459 (2011).
[12]V. L. Deringer, M. Lumeij, R. P. Stoffel, and R. Dronskowski, Chem. Mater. **25**, 2220 (2013).
[13]A. F. Voter, in *Radiation Effects in Solids*, edited by K. E. Sickafus and E. A. Kotomin (Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2005).
[14]K. Kreuer, Solid State Ionics **97**, 1 (1997).
[15]K. Kreuer, Solid State Ionics **136-137**, 149 (2000).
[16]K. D. Kreuer, Annu. Rev. Mater. Res. **33**, 333 (2003).
[17]M. A. Gomez, M. Chunduru, L. Chigweshe, L. Foster, S. J. Fensin, K. M. Fletcher, and L. E. Fernandez, J. Chem. Phys. **132**, 214709 (2010).
[18]M. A. Gomez, M. Chunduru, L. Chigweshe, and K. M. Fletcher, J. Chem. Phys. **133**, 064701 (2010).
[19]M. A. Gomez, D. Shepardson, L. T. Nguyen, and T. Kehinde, Solid State Ionics **213**, 8 (2011).
[20]M. A. Gomez and F.-J. Liu, Solid State Ionics **252**, 40 (2013).
[21]Y. Liu, M. Yoshino, K. Tatsumi, I. Tanaka, M. Morinaga, and H. Adachi, Mater. Trans. **46**, 1106 (2005).
[22]M. E. Björketun, P. G. Sundell, G. Wahnström, and D. Engberg, Solid State Ionics **176**, 3035 (2005).
[23]M. E. Björketun, P. G. Sundell, and G. Wahnström, Phys. Rev. B **76**, 054307 (2007).
[24]A. C. T. van Duin, B. V. Merinov, S. S. Han, C. O. Dorso, and W. A. Goddard III, J. Phys. Chem. A **112**, 11414 (2008).
[25]B. Merinov and W. Goddard III, J. Chem. Phys. **130**, 194707 (2009).
[26]See http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html for Chap. 11 of Grinstead and Snell.[3]
[27]I. Wolfram Research, *Mathematica version 10.0*, Wolfram Research, Inc., Champaign, Illinois, 2014.
[28]G. Kresse, "*Ab initio* molekular dynamik für flüssige metalle," Ph.D. thesis (Technische Universität at Wien, 1993).
[29]G. Kresse and J. Hafner, Phys. Rev. B **47**, RC558 (1993).
[30]G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[31]G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[32]G. Kresse and J. Joubert, Phys. Rev. B **59**, 1758 (1999).
[33]The labels on these PAW method files within the VASP library are Ba_sv, Zr_sv, Y_sv, O, and H.
[34]G. Henkelman, B. P. Uberuaga, and H. Jonsson, J. Chem. Phys. **113**, 9901 (2000).
[35]See http://theory.cm.utexas.edu/vtsttools/dynmat/ for VTST tools.
[36]H. G. Bohn and T. Schober, J. Am. Ceram. Soc. **83**, 768 (2000).
[37]T. Yajima, H. Suzuki, T. Yogo, and H. Iwahara, Solid State Ionics **51**, 101 (1992).
[38]K. Momma and F. Izumi, J. Appl. Crystallogr. **44**, 1272–1276 (2011).